Using magnetic resonance to image the pharynx during Arabic speech: Static and dynamic aspects

Ryan K. Shosted[1], *Bradley P. Sutton*[2], *Abbas Benmamoun*[1]

[1]Department of Linguistics, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[2]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
rshosted@illinois.edu, bsutton@illinois.edu, benmamou@illinois.edu

Abstract

Magnetic resonance imaging has been applied only recently to the study of Arabic speech production. Arabic has a relatively large number of sounds produced with constrictions in the pharynx, a part of the vocal anatomy well-suited to investigation using MRI. We show that static 3D MRI techniques can be useful in distinguishing the pharyngeal sounds of Arabic and that average pixel intensity in MR images can be used to track pharyngeal articulations as a function of time.

Index Terms: magnetic resonance, pharynx, pharyngeals, Arabic

1. Introduction

"[T]here is a considerable amount of guesswork involved in estimating the true shape and lateral dimensions of the pharynx cavity during back tongue articulation" [1, p. 97]. With the advent and wide dissemination of magnetic resonance (MR) technology, this guesswork is increasingly limited [2]. MR imaging allows researchers to examine multiple transverse / axial slices of the pharynx, particularly during static articulations of the same sound [3]. Challenges remain for dynamic (real-time) MRI data acquisition in terms of both temporal and spatial resolution, but improved techniques are developing rapidly [2, 4].

While the dimensions of the pharynx should arguably be studied in greater depth during the production of speech sounds in all languages, for languages where sounds are posited to include a primary or secondary constriction in the pharynx, the case is of particular interest. Arabic is such a language [5]. In this study, we will apply current methods in static and dynamic MRI to the production of the Arabic pharyngeal consonants /ħ ʕ/, the pharyngealized fricative /sˤ/, the vowels /a i u/ preceding pharyngeal and plain consonants, and the glottal fricative /h/.

The use of MRI in the study of Arabic has so far been limited to investigating vertical displacement of the larynx and pharyngeal width, measured in the mid-sagittal plane [6]. We will show that it is possible to establish critical differences between the relevant Arabic speech sounds using an MR-estimated pharyngeal area function based on transverse slices. We will also show that it is possible to track changes in the magnitude and location of pharyngeal cavity constriction as a function of time, using dynamic MRI. This advance could bring an important perspective to the study of pharyngealization as a time-varying, suprasegmental phenomenon in Arabic and other languages.

2. Methods

2.1. Speakers

One male speaker of Moroccan Arabic (MA1) and one male speaker of Jordanian Arabic (JA1) participated in the experiment. Data collection is ongoing for an additional ten speakers of Levantine and Gulf Arabic. We intend to investigate differences between these dialects in the future.

2.2. Procedures

For static MRI, each speaker produced sustained vowel and fricative productions for approximately 20 sec while lying supine inside a Magnetom Trio 3T scanner in the Biomedical Imaging Center at the Beckman Institute for Advanced Science and Technology. The scanning technique, known as a 3D Flash VIBE (Volume Interpolated Breath-hold Examination), allows for 3D acquisitions (here, using a transverse orientation) with a relatively high resolution (volume = $512 \times 416 \times 60$ voxels; resolution of 0.53 mm in plane and 2 mm resolution in the anterior-posterior direction) [7]. This allows us to estimate the size and shape of the pharyngeal cavity in inferior–superior steps of 2 mm during the production of vowels and fricatives. The 3D FLASH VIBE scans were acquired in approximately 20 sec. Static mid-sagittal scans were also acquired. A Turbo Spin Echo (TSE) sequence (20 sec in duration) was used for these scans with a bidimensional resolution of 0.78 mm/pixel in both directions. Subjects were trained before the session to maintain stable, sustained articulations of speech sounds for 20 sec.

For dynamic scans, speakers produced words with and without pharyngealized consonants embedded in the carrier phrase [iktubu ___ sɪt mərət] ‘I wrote ___ six times’.[1] Speakers repeated the sentence about 10 times throughout the scan. For these scans we used a multishot FLASH spiral design. The slice prescription consists of a single 6.5 mm thick midsagittal slice with 64 $\times$ 64 matrix and 240 mm field of view (FOV), resulting in a spatial resolution of 1.875 $\times$ 1.875 mm in the imaging plane. Sequence parameters for a 6-shot acquisition are TR (repetition time) 6.7 msec, TE (echo time) alternating between 0.9 and 1.4 msec (in order to obtain a field map for correcting the images), readout bandwidth of 200 kHz, and 1000 repetitions. Total scan duration was about 37 sec, with an average 63.3 msec per image or 15.8 frames/sec. Field map estimation, spatial encoding reduction, precise timing calculation, audio alignment, and simulation of susceptibility artifacts and correction were implemented as in [4]. Using a sliding window reconstruction, the reconstructed frame rate was 30 frames/sec. Subjects wore an MR-compatible headset with an attached optical microphone. The micropohone’s noise cancellation software was used to adequately sample and reduce the scanner’s background noise. While the resulting acoustic signals were not suitable for spectral analysis, it was possible to align events in the MR images with the help of the sound pressure signal and a spectrogram readout showing up to 2 kHz (Figure 4). All test material for recordings were presented on screen to the subjects using a Brainlogics MR digital projector.

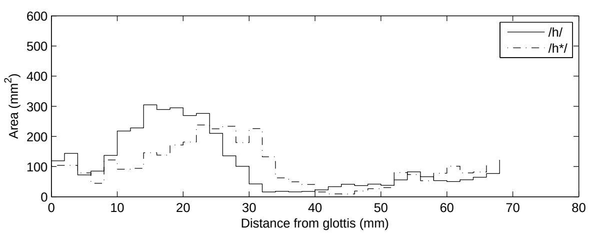

Figure 1: *Pharyngeal area functions of Arabic fricatives (JA1):* h* = /ħ/ (dot-dash).

2.3. Analysis

2.3.1. Static MRI

For static MR images, the contours of the pharynx were delineated on each transverse slice, proceeding upward from the most constricted portion of the glottis (just above the tracheal opening) to the last slice below the tip of the uvula. This included any volume between the tongue root and the epiglottis (if coupled to the main pharyngeal cavity), as well as the piriform sinuses. The level set algorithm [8] used in MIPAV 5.4.1 [9] was used to delineate these contours. Contour points were exported as XML and the area of each contour was then measured with the *polyarea* function in Matlab 7.10. The transverse pharyngeal slices associated with /ħ/ (orthographically ḥ) for JA1 accompany the electronic version of this article as `slices.eps`.[2] The pharyngeal area functions offer this data in summary fashion (Figures 1–3).

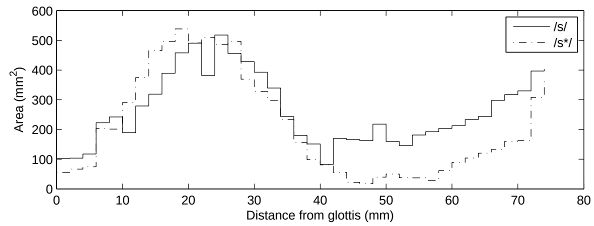

Figure 2: *Pharyngeal area functions of Arabic fricatives (JA1):* s* = /sˤ/ (dot-dash).

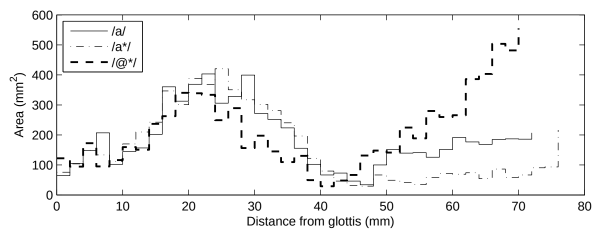

Figure 3: *Pharyngeal area functions of Arabic speech sounds (JA1):* a* = /aˤ/ (dot-dash); @* = /ʕ/ (dash).

2.3.2. Dynamic MRI

The overal intensity of the dynamic scans varies from image to image due to the alternating echo times used in the acquisition. Alternating echo times were needed to dynamically measure the field map and allow for the correction of image distortions [4]. This high frequency image intensity variation made it necessary to post-process the images for increased contrast. This was done by calculating the average pixel intensity m of each image I and transforming the image to a high-contrast image I_n using the formula $I_n = 1/(1 + m/(I + 2.2 \times 10^{-16}))^{10}$ (resulting images can be seen in Figures 4 and 5).

It has been argued that “kinematics of constriction formation and release can be estimated directly from regional pixel intensity variation in MR [i]mage sequences” [2]. This technique is in fact recommended, based on the fact that “anatomical delineation—or tissue tracking—techniques. . . are susceptible to segmetation error and are typically labor- and computationally intensive” (ibid.).

[1] In Moroccan Arabic, the first word of the carrier phrase is [ktəbu] and the second word is [sət].

[2] Throughout this paper, electronic attachments will be referenced using `this typeface`.

For the dynamic MR images, we selected five rectangular regions of interest corresponding approximately to the larynx, the hypopharynx, the region near the epiglottis, the oropharynx, and the nasopharynx. Each rectangle circumscribed its section of the larynx / pharynx from a mid-sagittal perspective, with dimensions of approximately 19 × 47 mm (`schema.eps`).

The pixel intensity within each rectangle was averaged. The time-aligned average pixel intensity (API) signal for the epiglottis, the audio spectrogram, and MR images at six different time points are given in Figure 4. An alternating echo time was used to jointly estimate the magnetic field map and image, producing oscillations from time point to time point. The lower-frequency variations in the signal show that a dynamic measure of constriction in the pharynx can be captured by the dynamic sequence. The utterance *iktubu rabaḥ sit marrat* is repeated three times in Figure 4.

3. Results

3.1. Static MRI

Because speech is inherently dynamic, the results of any static MR investigation of speech sounds will be limited. Nevertheless, it has been shown elsewhere that valuable articulatory insights can be gained from prolonged articulations like the ones in the current study [3, 10]. In this study, the area functions estimated using transverse slices of the pharynx point to important differences between the pharyngeal and pharyngealized (also known as 'emphatic') sounds of Arabic. For example, Figure 2 illustrates that the pharyngeal difference between a plain /s/ and pharyngealized / emphatic /sˁ/ occurs in the upper region of the pharynx. Inspection of corresponding sagittal images confirms this (`pls.jpeg, phs.jpeg`). Others have also claimed that pharyngealized coronals manifest a pharyngeal constriction higher than that of pharyngeal consonants [11, p. 124].

The pharyngeal differences between the glottal fricative /h/ and the pharyngeal fricative /ħ/ are apparent in Figure 1. With respect to the glottal fricative, the pharyngeal fricative manifests a constriction about 10–20 mm above the glottis. The area functions of the vowel /a/ produced before a plain coronal [d] and the same vowel produced before a pharyngealized / emphatic [dˁ] (Figure 3) show differences mainly in the upper pharynx. The area function of the pharyngealized vowel diverges from that of the plain vowel at about 50 mm above the glottis, where the pharyngealized variant becomes much more constricted. Besides a relatively small divergence in area function starting at about 25 mm above the glottis (where it is is more constricted), the pharyngeal approximant /ʕ/ differs from both variants of /a/ by manifesting a dramatic pharyngeal expansion beginning at about 50 mm above the glottis. This is confirmed in a mid-sagittal image of

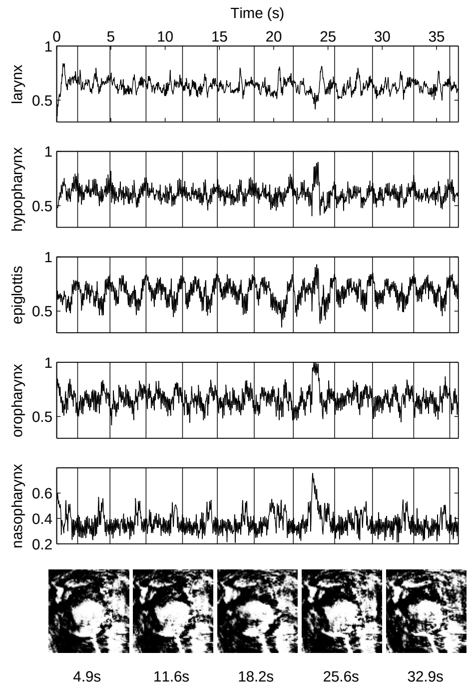

Figure 5: *Time-varying API at various regions of the pharynx. Vertical lines indicate occurrence of* /ħ/, *as judged by spectrogram inspection. MR images are shown for some of these time points.*

/ ʕ/ (`ayn.jpeg`) which reveals the epiglottis to approximate the pharyngeal wall quite closely while the tongue root is projected forward. The role of the epiglottis as an articulator has been contested [12]. Collected from a sufficient number of speakers, MRI data could be useful in further investigating this topic.

3.2. Dynamic MRI

Principal components analysis (PCA) of the time-varying average pixel intensity (API) for the laryngeal, hypopharyngeal, epiglottal, oropharyngeal, and nasopharyngeal regions revealed that two PCs account for on average about 85% of the variance contained in these signals (across all the utterances). While the first PC (PC1) has positive loadings / coefficients for API in all regions, the regions split into two groups along the axis of PC2, which has positive loadings for the oropharyngeal and nasopharyngeal regions and negative loadings for the laryngeal, hypopharyngeal, and epiglottal regions (a graphical representation of the PCA for *rabaḥ* is provided in `rabah_pca.eps`). This suggests that the dimensionality of the data may be reduced, i. e., that only an upper and lower portion of the pharynx are of practical value in future analyses of this kind of data. These trends were observed across different consonant types, including pharyngeal / non-pharyngeal (e. g., *rabab* / *rabaḥ*) and pharyngealized / non-pharyngealized (e. g., *rakid* / *rakiḍ*).

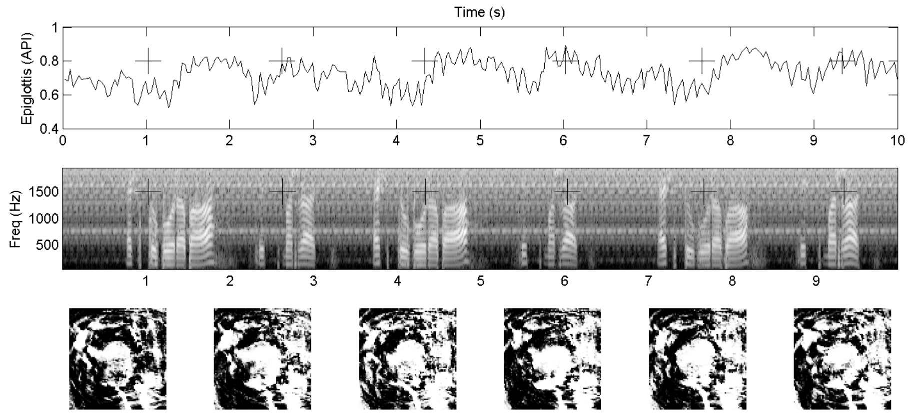

Figure 4: *Three repetitions of the utterance* iktubu rabaḥ sit marrat *(JA1), including time-varying API in the epiglottal region (top panel), audio spectrogram (middle panel), and MR images corresponding to the time points highlighted by '+' marks in the top two panels. These correspond to the release of [t] and the initiation of [u] in iktubu; the initiation of the 1st [ə] in marrat; the [r] in rabaḥ; the 2nd [ə] in marrat; the 2nd [u] of iktubu; and, again, the 2nd [ə] in marrat.*

Time-varying API, as illustrated in the top five panels of Figure 5, manifests considerable noise. However, it is possible to discern time-aligned peaks in the region of the epiglottis. The vertical bars in Figure 5 correspond to the production of /ħ/ as observed in the audio signal. While there are indeed other, non-aligned API peaks in the epiglottal region, a number of peaks do reliably cooccur with /ħ/. While there is room for greater precision in our measures, we note that the area function in Figure 1 suggests a constriction fairly deep in the pharynx. Thus, we might have expected clearer evidence of coincident peaks in the hypopharyngeal region for /ħ/.

4. Conclusions

The observations in this study are preliminary but have demonstrated the feasibility of using static transverse MRI to capture articulatory differences between pharyngeal, pharyngealized, and 'plain' speech sounds of Arabic. The study has also shown that it is feasible to collect dynamic MR images and correlate pixel intensity with pharyngeal articulation in at least some regions of the pharynx. Improvements in spatial resolution may allow us to reliably track the development and release of pharyngeal constriction as a function of time and to examine coordinative patterns.

5. Acknowledgements

We thank Paula Martins and colleagues at the University of Aveiro and the University of Coimbra for sharing their implementation of the VIBE sequence and for generously piloting static scans with the first author in Fall 2011. We acknowledge the support and assistance of Tracey Wszalek, Ryan Larsen, Wenmiao Lu, Nancy Dodge, and Holly Tracy at the Biomedical Imaging Center of the Beckman Institute for Advanced Science and Technology.

6. References

[1] Fant, G. "Acoustic theory of speech production", Mouton, 1960.

[2] Narayanan, S., Bresch, E., Ghosh, P., Goldstein, L., Katsamanis, Y., Kim, A., Lammert, A., Proctor, M., Ramanarayanan, V. and Zhu, Y. "A multimodal real-time MRI articulatory corpus for speech research", in Proc. Interspeech, 837–840, 2011.

[3] Demolin, D., Delvaux, V., Metens, T. and Soquet, A. "Determination of velum opening for French nasal vowels by magnetic resonance imaging", J. Voice, 17(2):1–14, 2003.

[4] Sutton, B. P., Conway, C. A., Bae, Y., Seethamraju, R. and Kuehn, D. P. "Faster dynamic imaging of speech with field inhomogeneity corrected spiral fast low angle shot (FLASH) at 3 T", J. Magn. Reson. Im. 32:1228–1237, 2010.

[5] Hassan, Z. M. and Heselwood, B. [Eds]. "Instrumental studies in Arabic phonetics", John Benjamins, 2011.

[6] Shar, S. and Ingram, J. "MRI investigation of Arabic gutturals", in W. -S. Lee and E. Zee [Eds], Proc. Intl. Cong. Phon. Sci. XVII, 1802–1805, City University of Hong Kong, 2011.

[7] Martins, P., Carbone, I., Pinto, A., Silva, A. and Teixeira, A. "European Portuguese MRI based speech production studies", Speech Comm., 50:925–952, 2008.

[8] Sethian, J. A. "Level set methods and fast marching methods", Cambridge University Press, 1999.

[9] McAuliffe, M., Lalonde, F., McGarry, D., Csaky, K. and Trus, B. "Medical image processing, analysis and visualization in clinical resarch" Proc. IEEE Symp. on Comp.-Based. Med. Sys., 381–386, 2001.

[10] Proctor, M. I., Shadle, C. H. and Iskarous, K. "Pharyngeal articulation in the production of voiced and voiceless fricatives", J. Acoust. Soc. Am. 127(3):1507–1518, 2011.

[11] Heselwood, B. and Al-Tamimi, F. "A study of the laryngeal and pharyngeal consonants in Jordanian Arabic using nasoendoscopy, videofluoroscopy and spectrography", in [5], 101–127, 2011.

[12] Esling, J. H. "The IPA categories 'pharyngeal' and 'epiglottal': Laryngoscopic observations of pharyngeal articulations and larynx height", Lang. Speech 42:349–372, 1999.

[11] Esling, J. H. "Pharyngeal consonants and the aryepiglottic sphincter", J. Intl. Phon. Assoc., 26:65–88, 1996.

[12] Esling, J. H. "Laryngoscopic analysis of pharyngeal articulations and larynx-height voice quality settings", Proc. 5th Intl. Conf. on Spoken Lang. Proc., ASSTA, 543–546, 1998.